2Б

УДК 681.324

С.А. Бутенков
Таганрогский технологический институт Южного федерального университета,
Таганрог, Россия
saab@tsure.ru

Математические модели анализа многомерных данных экологического мониторинга на основе теории информационной грануляции

Предлагаемый интеллектуальный подход к решению задачи сбора, хранения, передачи, обработки и анализа многомерных данных, экологического мониторинга средствами дистанционного зондирования основывается на систематическом введении методов мягких вычислений и имеющий целью повышение коэффициента интеллектуальности систем мониторинга. В результате использования основных идей теории информационной грануляции L. Zadeh вводится математическая модель многомерных данных, позволяющая решать все указанные задачи мониторинга с помощью методов мягких вычислений – вычислений над образами, или формами.

Введение

Развитие автоматической интерпретации многомерных данных и ГИС-технологий в области экологического мониторинга связано с тем, что человеку с его ограниченным трехмерным пространственным воображением сложно, а в большинстве случаев невозможно, анализировать и давать обобщенные оценки многомерным объектам [1], [2]. Существуют различные подходы к организации многомерных данных экологического мониторинга: по характеру решаемых задач, по уровням организации, по природным средам, за которыми ведутся наблюдения и т.д.

Мониторинг охватывает весь широкий спектр анализа наблюдений за меняющейся абиотической составляющей биосферы и ответной реакцией экосистем на эти изменения, включая как геофизические, так и биологические аспекты, что определяет широкий спектр методов и приемов исследований, используемых при его осуществлении.

Основой процесса дистанционного наблюдения за земной поверхностью [3], [4] является визуальное представление многомерных данных (изображения поверхности) как двумерной функции, возрастающей в соответствии с откликом различных областей на поверхности Земли на внешние источники энергии (луч локатора или Солнце для пассивного зондирования).

Основным качеством данных экологического мониторинга [2-4] является многомерность (цветность, мультиспектральность и т.д.). Другим практически важным качеством подобных данных является плохая определенность их геометрической формы. Общим термином, объединяющим возможные виды искажений информации, в том числе и изображений, является «неопределенность» (vagueness) [5]. Важнейшим понятием в рассматриваемой области представления неопределенных и низкокачественных изображений является понятие «неопределенного объекта» (vague region) [6], введенное в

ГИС. Неопределенные объекты чаще всего описываются своей границей. Объекты с неопределенной границей можно разделить на следующие классы:

– объект имеет границу, но ее форма неизвестна, либо не может быть определена путем измерений;

– объекты, которые не имеют определенной границы, или имеют границу, установленную условно. К ним относятся и нечеткие объекты, являющиеся предметом изучения нечеткой геометрии;

– объекты, не имеющие границы по определению (температура и влажность воздуха и т.п.). Подобные объекты можно представить как множества с определенным на них отношением частичного порядка, на котором вводится весовая функция $\varphi(x, y) = z$. Классы эквивалентности множества значений скалярного поля, для которых весовая функция принимает фиксированные значения $z_i, i = 1,...,n$, называются областями уровня и обычно служат объектами представления и анализа такого рода данных.

Практически важным классом неопределенных объектов являются природные образования, подобные морям, озерам и т.п. В водной среде понятие границы, формы объекта быстро изменяется в зависимости от действия множества природных факторов [7]. Эти свойства объекта изучения определяют спектр задач, решаемых в данной работе применительно к системам биологического мониторинга [7].

Цель работы

Целью работы является решение задач, связанных с развитием новых методов интеллектуальной обработки многомерных данных на основе теории информационной грануляции (ТИГ) [8] и перцептуального подхода к представлению многомерных данных [9].

Нужен подход, позволяющий на единой основе хранить, обрабатывать и анализировать такие данные независимо от размерности, обеспечивать регулярное представление плохо определенных данных, контролировать качество представления, компактифицировать [10]. Это направление объединяет методы анализа изображений и других многомерных данных в ГИС экологического мониторинга [11].

Некоторые результаты на этом пути были получены нами в работах по применению методов ТИГ к изображениям [9], [10], [12], [13], для нормализованных данных (приведенных к осям декартовой системы). Предложенный метод обеспечивает регулярное покрытие исходных неопределенных объектов (грануляцию данных) с помощью грассмановских гранул в декартовой системе координат [9]. Развитый метод позволяет достичь компактности, вычислительной регулярности, оптимизации степени сжатия данных и единообразие алгоритмов обработки для данных разной размерности в декартовых координатах [12-14]. Однако специфика использования данных дистанционного зондирования со спутников, летательных аппаратов (ЛА) и т.п. требует инвариантности к поворотам осей системы координат. Кроме того, некоторые виды данных представляются в системах координат, отличных от декартовой: например, цветные изображения представляют в цилиндрической и конической системах, пространственные неоднородности – в сферической и т.д. [13].

В качестве основной цели следует указать создание методологической базы для перехода от моделей многомерных данных в декартовых координатах к моделям в произвольных ортогональных системах координат и обобщение основных алгоритмов обработки и анализа многомерных данных на этот случай. Важной задачей является также исследование вопроса о «прозрачности» новых математических моделей для пользователей на основе так называемых моделей «стеклянного ящика» [14].

Постановка задачи

Теоретической основой для развиваемого в наших работах подхода является геометрическое истолкование задач и методов гранулирования многомерной информации [4], [5], поэтому в качестве математической основы работы выбрана геометрия грассмановых элементов, введенная F. Klein [15]. В работе основные результаты [15] обобщаются на модели в произвольных системах ортогональных координат, разрабатываются методики кластеризации гранул, основанные на использовании перцептуального подхода в интеллектуальной обработке информации [13], [14].

В качестве области приложений в работе рассматривается задача биомониторинга состояния поверхностных вод [2], [7] с использованием в качестве индикатора морского фитопланктона, оптическая плотность распределения которого наблюдается с низкого спутника или ЛА [3]. В связи с тем, что оптическим агентом фитопланктона является хлорофилл, полученные методы мониторинга могут использоваться для распространения на другие виды биоиндикаторов состояния среды, такие, как зеленые растения литорали и суши.

Дифференциальные модели неоднородности распределения фитопланктона

К числу важнейших закономерностей развития фитопланктона относится неравномерность пространственного распределения фитопланктона (пятнистость) [16]. Причиной этого явления служит еще недостаточно изученный механизм эктокринного регулирования роста фитопланктона. Математические модели используются для анализа динамики фитопланктонных популяций в предположении, что фитопланктоном выделяется биологически активный метаболит, концентрация которого влияет на скорость роста особей. Однако в силу того, что многие положения, лежащие в основе модели эктокринного регулирования, являются качественными, пятнистость является скорее качественной, чем количественной характеристикой [16]. При этом колебания плотности столь велики, что не могут быть объяснены случайными флуктуациями, и визуальная картина такова, что сравнительно небольшие площади высокой плотности («пятна», «облака») разделены пространствами с низкими плотностями, иногда не фиксируемыми стандартными методами наблюдений [2]. Особенно ярко это явление выражено в водоемах и районах океана, бедных биогенными элементами, где равномерное распределение фитопланктона не обеспечивало бы интенсивности первичного продуцирования, достаточной для развития высших трофических уровней [16]. Результаты численного моделирования по [16] приведены на рис. 1. Они демонстрируют образование устойчивых пятен, которые не разрушаются под воздействием течений и сохраняют свою форму и размеры.

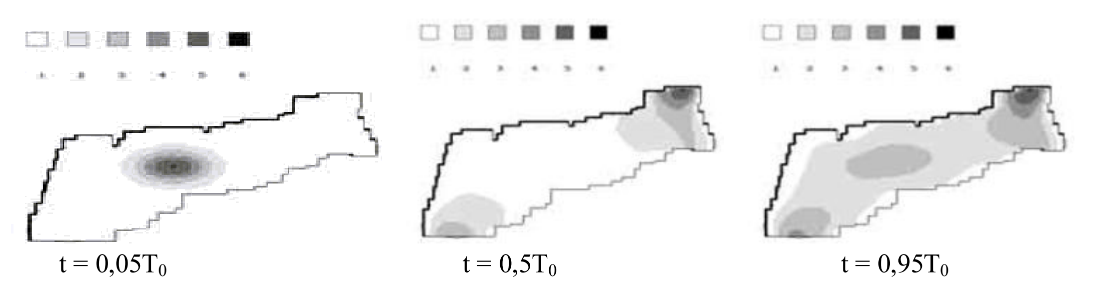


Рисунок 1 – Распределение фитопланктона в Таганрогском заливе по результатам моделирования

Отметим, что приведенные в [16] результаты моделирования получены с использованием большого объема вычислений и требуют времени порядка нескольких суток для одного прогноза при использовании двухъядерных персональных компьютеров. С точки зрения количественного анализа поведения популяций фитопланктона очевидно, что многие принципы, положенные в основу моделирования их роста, являются качественными [16], а с другой стороны, исчерпывающий анализ динамики дифференциальных моделей (например, разработанных в [16]) невозможен в силу отсутствия аналитических решений для таких моделей и необходимости значительных затрат времени при численном моделировании. Естественным в такой ситуации кажется применение гранулированных моделей динамики фитопланктона, основанных на идеях перцептуального подхода [8], [9] и мягких (качественных) вычислений [12], [13]. Согласно идеологии ТИГ такие методы на порядки снижают вычислительную емкость задач [8].

Для решения подобных задач необходимы математические модели распределения фитопланктона, позволяющие анализировать многомерные данные о наблюдаемой форме его популяций [4] на основе геометрических соображений [11]. Для подобных целей в ряде работ использовались методы информационной грануляции [12-14], однако они не учитывали целый ряд особенностей геометрии мониторинга с ЛА и спутников [3]. Введем основные положения, развивающие эти результаты на данную область применения.

Гранулированные модели многомерных данных (распределения фитопланктона)

Поскольку в ТИГ идеология пространственной грануляции только начинает разрабатываться [8], то здесь необходима разработка новых подходов.

Информационной гранулой называется подмножество универсума, на котором определено отношение сходства, неразличимости и т.п. [10]. Множество гранул, которое содержит все объекты универсума, называется *гранулированием универсума*. Подмножество $A \subseteq U$ называется *составной* (не элементарной) *гранулой*, если оно представляет собой объединение *атомарных гранул* [10].

Определив на плоскости проекции произвольной гранулы A, обозначаемые как $pr_x A$ и $pr_y A$, зададим *инкапсулирующую декартову гранулу* для произвольного A как $G^+ = pr_x A \times pr_y A$. Гранула G^+ является точной верхней гранью конечного множества всех гранул, содержащих A.

В ряде наших работ [9], [10] была использована алгебраическая модель декартовой гранулы, позволяющая компактно кодировать геометрическую информацию в виде матриц специального вида, которые F. Klein назвал «грассмановыми элементами» [15]:

$$g(x_1, x_2, x_3, y_1, y_2, y_3) = \begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix}. \quad (1)$$

На основе таких элементов F. Klein построил полную геометрическую теорию для плоскости и пространства [15]. Согласно [15], принципиальную роль для построения геометрии гранул (1) играет определитель (1), имеющий геометрический смысл суммы алгебраических площадей треугольников, построенных на опорных точках элемента (гранулы), как показано на рис. 2а).

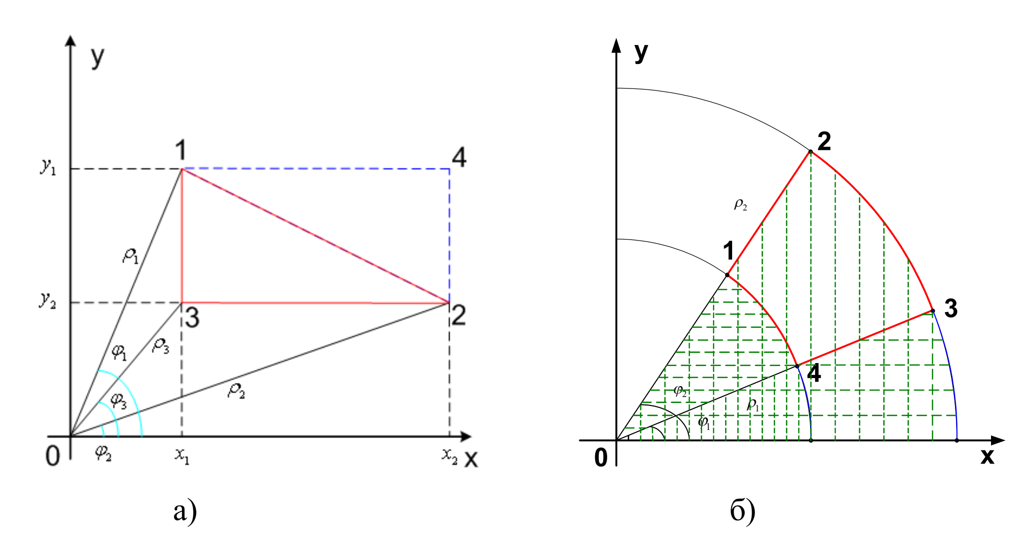


Рисунок 2 – Определение базового элемента для плоскости:
а) в декартовой системе координат (СК), б) в полярной СК

Согласно [12], [13], [15], из модели (1) можно получать геометрическую информацию о базовом элементе (в дальнейшем будем называть его гранулой). Так, высота и ширина гранулы находятся с помощью следующих миноров базового элемента:

$$h = \left\| \begin{matrix} y_1 & 1 \\ y_2 & 1 \end{matrix} \right\|, \qquad w = \left\| \begin{matrix} x_1 & 1 \\ x_2 & 1 \end{matrix} \right\|. \tag{2}$$

Геометрические характеристики гранулы – периметр, диагональ и площадь гранулы соответственно – выражаются через миноры базового элемента как

$$P = 2\left(\left\| \begin{matrix} y_1 & 1 \\ y_2 & 1 \end{matrix} \right\| + \left\| \begin{matrix} x_1 & 1 \\ x_2 & 1 \end{matrix} \right\| \right), \quad l = \sqrt{\left\| \begin{matrix} y_1 & 1 \\ y_2 & 1 \end{matrix} \right\|^2 + \left\| \begin{matrix} x_1 & 1 \\ x_2 & 1 \end{matrix} \right\|^2}, \quad S = \left\| \begin{matrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{matrix} \right\|. \tag{3}$$

Базовые модели гранул для различной размерности данных

Используя принцип инвариантного кодирования геометрической информации по [13], мы можем получить уравнения базовых элементов в различных ортогональных системах координат. В полярной системе координат гранула $1,2,3,4$ может быть определена с помощью площадей базовых секторов. По аналогии с (1) введем базовый алгебраический элемент в полярной системе, заданный предельными значениями полярных радиусов ρ_1 и ρ_2 и полярных углов φ_1 и φ_2:

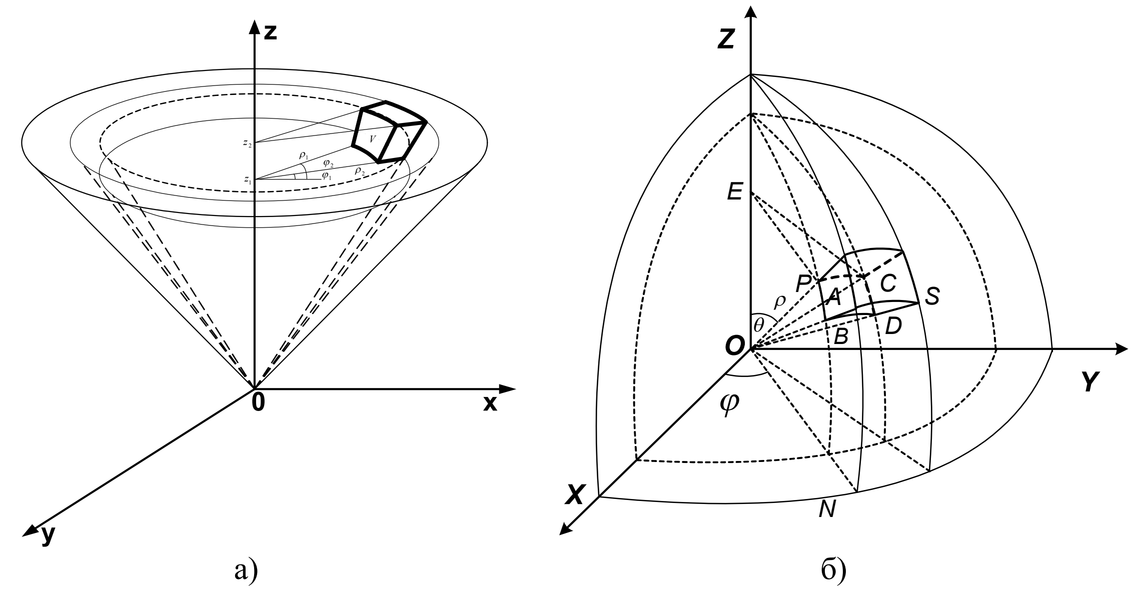
$$g^{Polar}(\rho_1, \rho_2, \varphi_1, \varphi_2) = \begin{pmatrix} \rho_1 & \rho_2 & 0 \\ \varphi_1\rho_2 & \varphi_1\rho_1 & 1 \\ \varphi_2\rho_2 & \varphi_2\rho_1 & 1 \end{pmatrix}. \tag{4}$$

Распространяя результаты предыдущего раздела на трехмерный случай, легко получить по аналогии с (4) уравнение базового элемента в цилиндрической системе

координат, заданного предельными значениями полярных радиусов ρ_1 и ρ_2, полярных углов φ_1 и φ_2 и высот z_1 и z_2 :ρ

$$g^{Cyl}\left(\rho_1,\rho_2,\varphi_1,\varphi_2,z_1,z_2\right)=\begin{pmatrix} z_1 & z_2 & 0 & 0 \\ \rho_2 & \rho_2 & \rho_1 & 0 \\ \varphi_2\rho_1 & \varphi_2\rho_1 & \varphi_2\rho_2 & 1 \\ \varphi_1\rho_1 & \varphi_1\rho_1 & \varphi_1\rho_2 & 1 \end{pmatrix}. \qquad (5)$$

Наиболее сложными случаями являются алгебраические описания гранул в виде части усеченного конуса и сферы (рис. 3).



а) б)

Рисунок 3 – Базовые элементы в конических (а) и сферических (б) координатах

По аналогии с (5), уравнение базового элемента в конической системе координат, заданного предельными значениями полярных радиусов ρ_1 и ρ_2, полярных углов φ_1 и φ_2 и высот z_1 и z_2, согласно рис. 3а) запишем в виде:

$$g^{Cone}\left(\rho_1,\rho_2,\varphi_1,\varphi_2,z_1,z_2\right)=\begin{pmatrix} z_1^3 & z_2^3 & 0 & 0 \\ \dfrac{\rho_2}{z_2^2} & \dfrac{\rho_2}{z_2^2} & \dfrac{\rho_1}{z_1^2} & 0 \\ \varphi_2\rho_1 & \varphi_2\rho_1 & \varphi_2\rho_2 & 1 \\ \varphi_1\rho_1 & \varphi_1\rho_1 & \varphi_1\rho_2 & 1 \end{pmatrix}. \qquad (6)$$

Используя (6), мы можем получить уравнение базового элемента в сферической системе координат, заданного предельными значениями полярных радиусов ρ_1 и ρ_2, полярных углов φ_1 и φ_2 и углов θ_1 и θ_2 согласно рис. 3б) в виде:

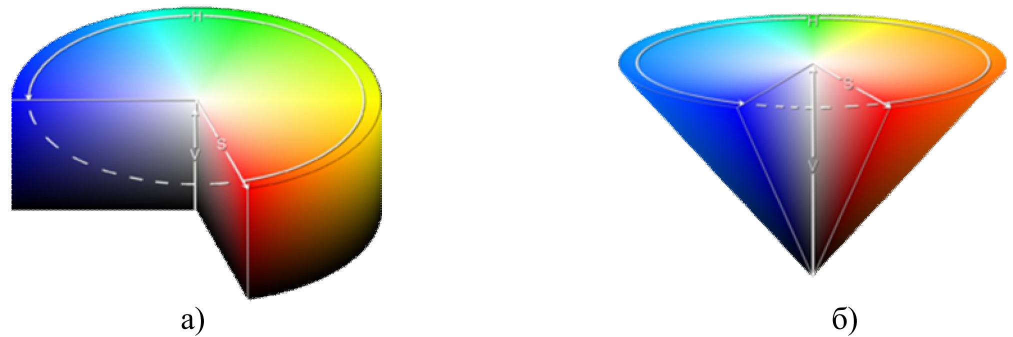
$$g^{S}\left(\rho_1,\rho_2,\varphi_1,\varphi_2,\theta_1,\theta_2\right)=\begin{pmatrix} \rho_2^2 & \rho_2^2 & \rho_1^2 & 0 \\ \cos\theta_1 & \cos\theta_2 & 0 & 0 \\ \varphi_1\rho_1 & \varphi_1\rho_1 & \varphi_1\rho_2 & 1 \\ \varphi_2\rho_1 & \varphi_2\rho_1 & \varphi_2\rho_2 & 1 \end{pmatrix}. \qquad (7)$$

Модели (4 – 7) могут использоваться для нахождения отдельных геометрических характеристик криволинейной гранулы путем вычисления определителей ее различных миноров. Практический пример, приводящий к необходимости использования модели (6), описан в следующем разделе.

Гранулирование в перцептуальном цветовом пространстве

Наиболее соответствующей цветовым перцепциям человека моделью восприятия цвета является модель HSV, координатами которой являются: Hue – цветовой тон, Saturation – насыщенность, Value – яркость.

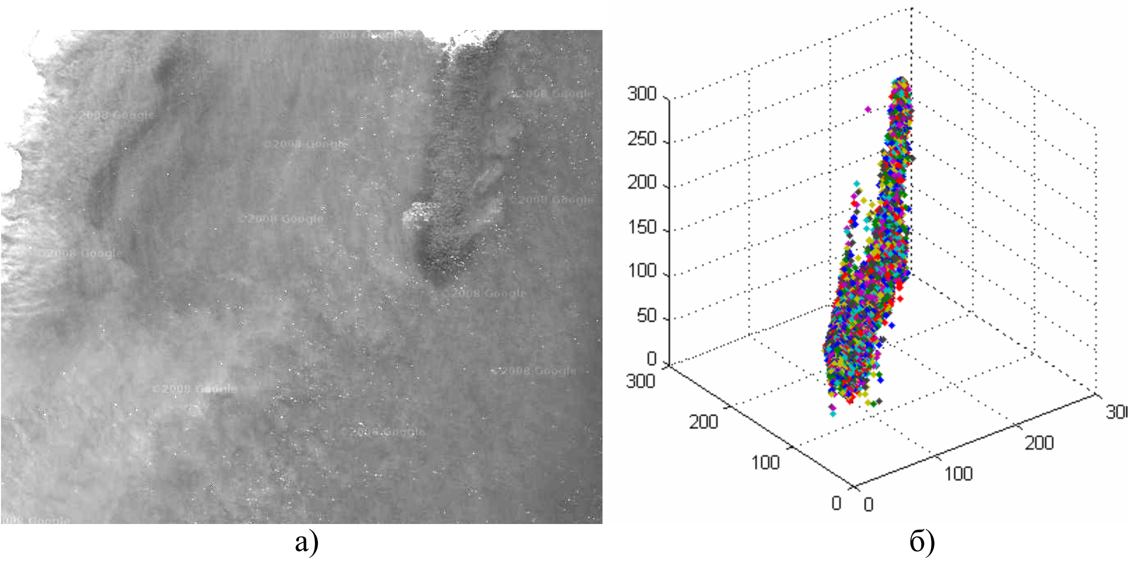
Простейшим способом отображения координат HSV в трёхмерное пространство является использование цилиндрической системы координат. Здесь координата H определяется полярным углом φ, S – радиус-вектором ρ, а V – координатой z (рис. 4а). Чаще применяется коническая цветовая модель, в которой значение радиуса зависит от высоты по оси z (рис. 4б).



а) б)

Рисунок 4 – Цилиндрическая (а) и коническая (б) модели цветового пространства

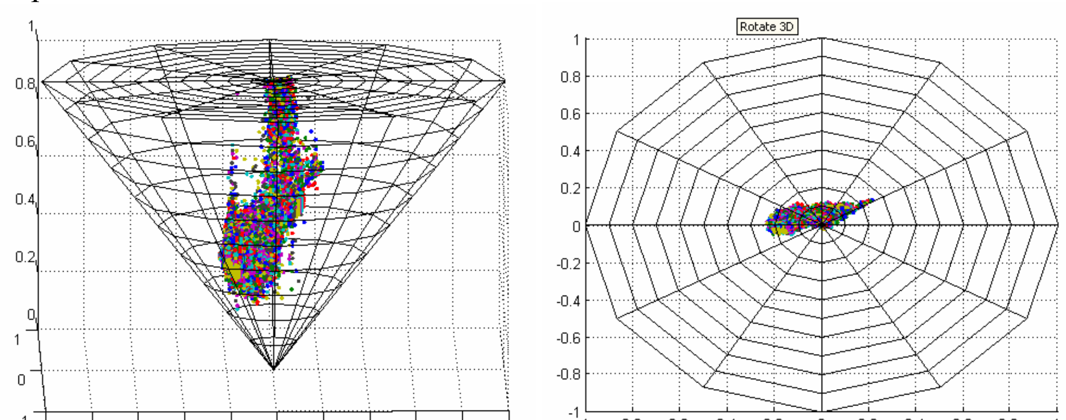
В цветовом пространстве HSV объект интереса представляется некоторым объединением гранул [8] в силу того, что его оттенок H и насыщенность S варьируют из-за естественного (или искусственного) изменения видимой цветности объекта [11]. Рис. 5 показывает результат грануляции по (6) гистограммы яркостей реального спутникового изображения.



а) б)

Рисунок 5 – Спутниковое изображение (а) поверхности моря в районе Барселоны и его гистограммы яркостей (б) в пространстве RGB (изображение GoogleEarth)

Рис. 6 показывает результат гранулирования изображения рис. 5а) в цветовом пространстве HSV.



а) б)

Рисунок 6 – Результаты гранулирования гистограммы в конической (а) и в полярной (б) системах координат

Рис. 6а) показывает полную гранулированную гистограмму яркостей, построенную по модели (6), а рис. 6б) – упрощенную гистограмму, полученную путем проецирования рис. 6а) на плоскость $V = \max$. Такое упрощение позволяет, практически не теряя информации, понизить размерность входных данных на единицу.

К представлению рис. 6 могут быть применены методы обработки многомерных данных, описанных в работах [12-14].

Выводы

Введенные в работе модели многомерных гранул позволяют унифицировать алгоритмы представления, передачи, обработки и анализа многомерных данных экологического мониторинга. Они в полной мере реализуют тот факт, что подавляющее большинство моделей (в том числе и «точных» дифференциальных) строятся на основе качественных рассуждений о механизме эктокринного регулирования фитопланктона.

Полученные общие результаты позволяют использовать методы гранулирования в широком спектре задач (путем выбора соответствующей модели (4 – 7)), в том числе и в задачах анализа изображений как данных пассивного оптического мониторинга состояния поверхностных вод моря.

В отличие от распространенных подходов к подобному представлению, основанному на выделении некоторых локальных особенностей или *признаков*, характеризующих объект [5], предложенное представление является регулярным (в смысле некорректно поставленных задач), поскольку основано на процессе инкапсуляции (покрытия) гранулами по L. Zadeh и использовании оценок площадей полученных декартовых гранул, что является аналогом интегрирования (в широком смысле этого слова) по подмножествам декартова произведения. Предложенные унифицированные алгебро-геометрические модели могут быть широко использованы при построении интеллектуальных систем анализа многомерных данных, «прозрачных» для пользователя или систем типа «стеклянного ящика» [14] в противовес гранулированным моделям типа «черного ящика» [5]. В результате они полностью реализуют все преимущества информационной грануляции, заявленные в [8].

Литература

1. Айвазян С.А. Классификация многомерных наблюдений / Айвазян С.А., Бежаева З.И., Староверов О.В. – М. : Статистика, 1974. – 240 с.
2. Красовский Г.Я. Аэрокосмический мониторинг поверхностных вод (практические аспекты) / Красовский Г.Я. – М. : Науч. совет по космич. исслед. для народн. хоз-ва МКС АН СССР. – 231 с.
3. Оптико-электронные системы экологического мониторинга природной среды : [сб. науч. тр. / ред. Рождествина В.Н.]. – М. : МВТУ им. Баумана. 2003. – 527 с.
4. Câmara G. What's in an Image / [Câmara G., Egenhofer M., Fonseca F. and Monteiro A.M.V. // Spatial Information Theory – A Theoretical Basis for GIS : [Proc. International Conference COSIT '01]. – Santa Barbara : (CA), 2001.
5. Rosenfeld A. From Image Analysis to Computer Vision / A. Rosenfeld // Computer Vision and Image Understanding. – 2001. – № 84. – P. 298-324.
6. Erwig M.Vague Regions / M. Erwig, M. Schneider // 5th Int. Symp. on Advances in Spatial Databases (SSD). – 1997. – P. 298-320.
7. Израэль Ю.А. Концепция мониторинга состояния биосферы / Ю.А. Израэль // Мониторинг состояния окружающей природной среды : [Тр. 1-го советско-английского симпозиума]. – Л. : Гидрометеоиздат, 1977. – С. 10-25.
8. Zadeh L. Towards a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic / L. Zadeh // Fuzzy Sets and Systems. – 1997. – № 90. – P. 111-127.
9. Бутенков С.А. Грануляция и инкапсуляция в системах эффективной обработки многомерной информации / С.А. Бутенков // Искусственный интеллект. – 2005. – № 4. – С. 106-115.
10. Бутенков С.А. Развитие парадигмы интеллектуального анализа многомерной информации применительно к теории информационной грануляции / С.А. Бутенков // Труды IV Международного научно-практического семинара «Интегрированные модели и мягкие вычисления в искусственном интеллекте», (Коломна, 28 – 30 мая 2007 г.) – 2007. – Т. 1. – С. 188-194.
11. Baltsavias E. Integration of Image Analysis and GIS / E. Baltsavias, M. Hahn // ISPS Comission IV Symposium "GIS – between Vision and Applications". – Stutgart (Germany), 1999. – P. 669-676.
12. Бутенков С.А. Анализ изображений в параллельных системах с помощью R-гистограмм для обобщенных геометрических моделей / С.А. Бутенков // Искусственный интеллект. – 2003. – № 3. – С. 200-208.
13. Бутенков С.А. Алгебраические модели в задачах интеллектуального анализа многомерных данных / С.А. Бутенков // Труды международной научно-технической конференции «Математическая теория систем 2009» (МТС-2009), (Москва, 26-30 января, 2009 г.). – 2009. – С. 93-101.
14. Butenkov S.A. Granular Computing in Computer Image Perception: basic issues and Glass Box models / S.A. Butenkov, V.V. Krivsha, S.Al. Dhouyani // Artificial Intelligence and applications "AIA 2006": [Proc. of IASTED Conf]. – Innsbruk (Austria), 2006. – P. 462-467.
15. Klein F. Elementarmathematik vom Hoheren Standpunkte Aus Erster Band / F. Klein. – Berlin : Verlag von Julius Springer, 1924.
16. Домбровский Ю.А. Пространственная и временная упорядоченность в экологических и биохимических системах / Ю.А. Домбровский, Г.С. Маркман. – Ростов н/Д., 1983. – 120 с.

С.А. Бутенков

Математичні моделі для аналізу багатовимірних даних на основі теорії інформаційної грануляції

Пропонується інтелектуальний підхід до розв'язання задачі збору, зберігання, передачі, обробки і аналізу багатовимірних даних, екологічного моніторингу засобами дистанційного зондування, заснований на систематичному введенні методів м'яких обчислень і такий, що має на меті підвищення коефіцієнта інтелектуальності систем моніторингу. В результаті використання основних ідей теорії інформаційної грануляції L. Zadeh вводиться математична модель багатовимірних даних, що дозволяє вирішувати всі вказані завдання моніторингу за допомогою методів м'яких обчислень – обчислень над образами, або формами.

S.A. Butenkov

Mathematical Models of Ecological Monitoring Multidimensional Data Analysis Based on the Information Granulation Theory

This paper deals with the perspectives of new approach to the common problem of multidimensional data of collection, transmission, processing and analysis. The main idea is related with the soft computing principles and aims the MIQ updating according the main propositions of Information Granulation Theory, proposed by L. Zadeh. The new kind of soft computing operations (computing with shapes) is proposed and used for satellite color images representation in compact form.

Статья поступила в редакцию 01.07.2009.